Analysis of Selection Patterns in Germline Stem Cell Genes (*Stonewall*, *Otefin*) in *Drosophila pseudoobscura*

◆Honors Thesis◆

Presented to the College of Arts and Sciences
Cornell University
In partial fulfillment of the requirements for the Biology Honors Program

By,
Lisebeth Forbes ’13

May 2013

Professor Charles F. Aquadro

Table of Contents

ABSTRACT

Previous experiments have shown two germline stem cell genes, *bam* and *bgcn*, to be under strong positive selection in *Drosophila melanogaster* and *Drosophila simulans* (Bauer DuMont et al. 2007). This prompted the question of whether the same pattern of selection observed in these two species was present in the germline stem cell genes of other *Drosophila* lineages? The Aquadro Lab has been sequencing many germline stem cell genes in *Drosophila* species, and the answer to this question so far has been that some lineages show strong positive selection and some do not. This observation led the Aquadro Lab to begin to test hypotheses about the driver – or drivers – of the positive selection in the germline stem cell genes across some *Drosophila* lineages. One hypothesis proposed by Bauer DuMont et al. (2007) is that coevolution with pathogens such as the reproductive parasite, *Wolbachia pipientis*, infecting the germline could be driving this observed selection. This project looked for signs of selection in the germline stem cell genes *stonewall* and *otefin*, two genes that have shown signs of positive selection in other *Drosophila* species, which the Aquadro Lab has previously tested. These two genes were sequenced in *Drosophila pseudoobscura*, a species of *Drosophila* that is not currently known to be infected with *Wolbachia*. This project shows that both genes do not show evidence of long term, repeated positive selection, but both genes do show evidence for a more recent selective sweep.

BACKGROUND

Germline stem cells (GSC's) in the gonads of *Drosophila* are crucial for the creation of eggs and sperm, in females and males, respectively. The correct formation of egg and sperm is essential to animal reproduction and therefore organismal fitness. The genes that control the formation and regulation of the GSC's are referred to here as

germline stem cell genes. This group of genes is vital to a species' survival because they are critical in maintaining the germline throughout an organism's life and thus its ability to reproduce. Therefore, it might be expected that evolutionary pressures would preserve these genes and keep them evolving under the neutral model with very strong constraint on amino acid variation within and between species.

Surprisingly, Civetta et al. (2006) and the Aquadro Lab (Bauer DuMont et al. 2007) discovered, independently, that one of the GSC genes, *bag of marbles* (*bam*), is under strong positive selection with a large number of nonsynonymous substitutions (59) among 442 codons between *Drosophila melanogaster* and *Drosophila simulans*, two closely related species (Bauer DuMont et al. 2007). Bauer DuMont et al. also discovered that a second GSC gene, which acts together with *bam* in germline stem cell differentiation, *benign gonial cell neoplasm* (*bgcn*), is also under strong positive selection in the two aforementioned species (Bauer DuMont et al. 2007). These results suggest rapid evolution at the protein level, which was surprising due to the importance of these genes to organismal fitness. These observations led to the following question: are the patterns of positive selection seen in *D. melanogaster* and *D. simulans* consistent across the other *Drosophila* lineages? Interestingly, the answer has been no. Some species have displayed strong positive selection in their GSC genes and some have not. The Aquadro Lab has been using data across different species of *Drosophila* to test hypotheses about the driver, or drivers, of the observed strong positive selection in some of the *Drosophila* lineages.

One hypothesis proposed by Bauer DuMont et al. 2007 was that coevolution with pathogens infecting the germline could be driving the observed positive selection.

One of these proposed pathogens was the maternally inherited bacterium *Wolbachia pipientis*. *Wolbachia* is a successful reproductive parasite. Its success largely comes from its ability to manipulate the reproductive success of females in favor of those infected by *Wolbachia*, its maternal inheritance (via the egg), as well as its ability to increase resistance in its host to some viral infections (e.g., Hedges et al. 2008; Teixeira, Ferreira, Ashburner 2008; and Werren et al. 2008). Evidence of infection with this bacterial endosymbiont has been observed in some of the *Drosophila* species but not in all of them (Mateos et al. 2006). To test this hypothesis that *Wolbachia* is a driver of positive selection across the *Drosophila* genus, the Aquadro Lab has been sequencing different GSC genes in many *Drosophila* species (some have shown evidence of *Wolbachia* infection and some have not) to see which show signs of positive selection. For this project, two GSC genes, *stonewall* (*stwl*) and *otefin* (*ote*), were sequenced in *D. pseudoobscura*, in which there has been no evidence of current *Wolbachia* infection (Figure 1). *Stwl* and *ote* were selected as the focus GSC genes of this project because both have shown evidence, in previous experiments conducted by the Aquadro lab, of long-term recurrent positive selection for amino acid diversification in *Drosophila* linages that have evidence for long-term infection with *Wolbachia* (*D. melanogaster* and *D. simulans*) (Flores et al. 2013 and Jae Choi, personal communication). In addition, *ote* was of special interest because it interacts with *bam* in the GSC (Figure 2).

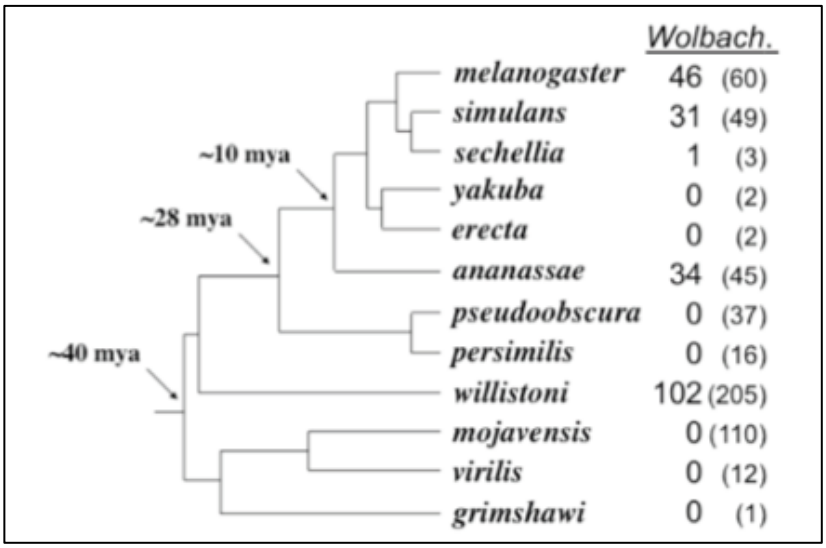

Figure 1 – Infectivity of *Drosophila* species with *Wolbachia*. The number on the left is the number of lines infected with *Wolbachia* of the lines tested, and the number in the parentheses is the number of lines that were tested for *Wolbachia*. Out of the 37 lines of *D. pseudoobscura* tested, zero were found to be infected with *Wolbachia* (Mateos et al. 2006, Watts et al. 2009, Montenegro et al. 2005, Haselkorn et al. 2009, and Flyendo: http://flyendo.arl.arizona.edu/).

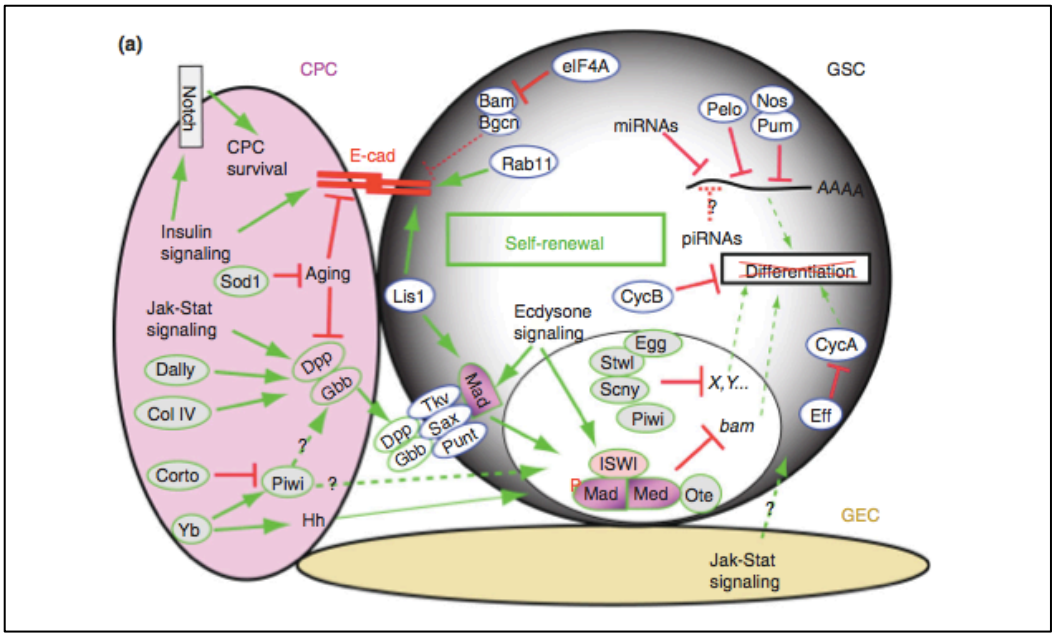

Figure 2 – Overview of extrinsic and intrinsic factors controlling GSC self-renewal. The pink ellipse represents cap cells ("CPC"), the grey circle represents germline stem cells ("GSC"), and the beige ellipse represents germline stem cell contacting escort cells ("GEC"). Solid green arrows indicate positive regulations whereas solid red arrows show inhibitory relationships. Broken green arrows show that regulations have been inhibited, and broken red arrows with question marks show potential relationships that have not been proven. (Xie, T. WIREs Dev Biol 2012. dio:10.1002/wdev.60)

Stwl is a GSC gene that encodes a chromatin-remodeling factor (Xie, T. 2012). *Stwl* is located on the right arm of the X chromosome in *D. pseudoobscura* at cytological band 28, and it is approximately 3,000 base pairs long. *Stwl* is located approximately 7.1 megabases from the centromere of the X chromosome and is therefore roughly halfway between the centromere and the tip of the 13 mega base long right arm of the X chromosome (Marygold et al. and the FlyBase Consortium 2013). The gene consists of two exons: a shorter first exon (approximately 100 base pairs) and a longer second exon (approximately 2900 base pairs). *Ote* encodes a nuclear lamin-binding protein, which, in turn, is a negative regulator of *bam* (Jiang, X et al. 2008 & Xie, T. 2012). *Ote* is located on the third chromosome in *D. pseudoobscura* at cytological band 73, and it is approximately 1,250 base pairs long. *Ote* is located approximately 11.4 megabases from one end of the third chromosome; it is within several third chromosome inversions and located next to the gene *Amylase I* (Marygold et al. and the FlyBase Consortium 2013).

Under the endosymbiont conflict hypothesis being tested by this project, *Wolbachia* acts as the driver of positive selection in at least some of the GSC genes. There are several results that would be consistent with this hypothesis. In species where there is evidence for long-term infection with *Wolbachia*, like *D. melanogaster* or *D. simulans* (Richardson et al. 2013; Jae Choi, personal communication), there should be signs of long-term, recurrent positive selection in the sequences of the GSC genes. Conversely, if a species is not infected with *Wolbachia*, then the sequence data of the GSC genes should fail to reject the hypotheses of neutrality in the population genetic tests that test for neutrality (and include strong selective constraint on protein evolution). However, there are also some other results that could occur and would also be consistent

with this hypothesis. If a species has no evidence of current infection with *Wolbachia*, but its sequence data rejects neutrality under the McDonald Kreitman Test (which looks for past, repeated positive selection) and fails to reject the hypothesis of neutrality under Tajima's Test and Fay and Wu's H (that looks for recent selective sweeps), then it is possible that the species was previously infected with *Wolbachia* and has only recently lost the infection. Also, if a species was previously not infected with *Wolbachia* and then only recently acquired the infection, then it would be possible to observe a McDonald Kreitman Test that fails to reject the hypothesis of neutrality and a Tajima's test and Fay and Wu's H that does reject the hypothesis of neutrality in a direction consistent with a recent selective sweep of new adaptive mutation in the gene. Also, it is always possible that if a gene shows evidence of a recent selective sweep, that there is a gene to which the tested gene is linked that is being selected for and not the tested gene itself.

MATERIALS AND METHODS

Fly Stocks and DNA lines

Thirty-five lines of *D. pseudoobscura* that had each originated from a single wild-caught female fly were used for the final analysis of *stwl*, and twenty-five lines were used for the final analysis of *ote*. The sources of these flies were Mesa Verde National Park (MV), Kaibab National Forest (KB), Bosque Del Apache National Wildlife Refuge (BMC), and Apple Hill, California (AH). In the final analysis of the data the break down of flies from each population were as follows: eight AH, nine MBC, seven KB, and eleven MV for *stwl*; ten BMC, six KB, and nine MV for *ote*.

Sequencing

Genomic DNA was extracted from adult flies from these lines using Purgene Core Kit A DNA Isolation kits (Qiagen). Polymerase Chain Reaction (PCR) primers

were made and used to amplify each gene region. Both flanking and internal sequencing primers were used for each gene or gene segment to achieve coverage of both strands of DNA. Sanger sequencing was performed by the Cornell University Genomics Core DNA Sequencing Facility (http://cores.lifesciences.cornell.edu/brcinfo/?f=1) using ABI Chemistry and 3730XL DNA Analyzers. The sequences obtained from the Core Sequencing Facility were edited and assembled in Sequencer 5.0.1 (Gene Codes). Once edited, the sequences were aligned in MEGA 5 using the muscle aligner (Tamura et al. 2007).

Sequencing was attempted on 30 lines of *D. pseudoobscura* for each gene. For each gene sequenced, between 20 and 28 lines were successfully sequenced. Due to difficultly in obtaining quality sequencing results for some lines, supplemental sequences, provided by graduate student Jae Choi were added to the final analyses of the two genes for a total of 35 lines for *stwl* and 25 lines for *ote* (See **Appendix C** for more details).

Polymorphism and Divergence Analysis

DnaSP 5.10.1 (Librado and Rozas 2009) was used to perform population genetics tests on the aligned sequence data. For all tests, *D. miranda* (sequence obtained from M. Noor, personal communication) was used as the outgroup. In order to detect signs of long-term, recurrent selection, the McDonald Kreitman Test was performed (McDonald and Kreitman 1991). In addition, to look for an excess (or deficiency) of rare variants, which could indicate a recent selective sweep (or balancing selection), Tajima's D and Fay and Wu's H were calculated (Tajima 1998; Fay and Wu 2000). Polymorphism and divergence data was also collected using DnaSP. P-values for each test statistic were obtained using the neutral coalescent simulator with recombination in

DnaSP unless otherwise noted. Levels of recombination were estimated using DnaSP and used, where noted, when calculating of P-values.

RESULTS

To measure the levels of polymorphism in both genes, two measures of nucleotide diversity, π and θ_w, were calculated. The test statistic π is the number of pairwise differences observed per nucleotide for the gene region (and provides one estimate of the population parameter, $4N_e\mu$), and θ_w estimates $4N_e\mu$ from the number of segregating sites in the region sequenced. For *stwl* and *ote*, the synonymous polymorphism, π(s), was 0.00175 and 0.01524, respectively (**Table 1**). For comparison, the average synonymous polymorphism for 100 other genes among 14 lines of *D. pseudoobscura* was 0.014 (Jensen and Bachtrog 2011). The nonsynonymous polymorphism, π(a), was 0.00086 and 0.00193 for *stwl* and *ote*, respectively. The average π(a) for 100 other genes of *D. pseudoobscura* was 0.0011 (Jensen and Bachtrog 2011). The θ_w for *stwl* and *ote* was 0.00708 and 0.01323, respectively, and the average θ_w for 100 genes of *D. pseudoobscura* was 0.019 (Jensen and Bachtrog 2011). The raw polymorphism data for *stwl* and *ote* are provided in **Appendix A** and **Appendix B**, respectively.

Table 1 – Estimates of nucleotide polymorphism in *D. pseudoobscura* and divergence to *D. miranda* for *stwl* and *ote*.

	Stwl	*Ote*
Nonsynonymous Polymorphism π(a)	0.00086	0.00193
Synonymous Polymorphism π(s)	0.00175	0.01524
π(a) / π(s) Ratio	0.487	0.126
Nonsynonymous Divergence k(a)	0.01829	0.01847
Synonymous Divergence k(s)	0.04042	0.04321
k(a) / k(s) Ratio	0.446	0.420
# of Segregating Sites	21	38
# of Nucleotide Sites in Region	3009	1248
θ_w	0.00708	0.01323

Tajima's D tests for an excess or deficiency of rare variants versus intermediate frequency alleles. The Tajima's D's for *stwl* and *ote* were negative indicating an excess of rare variants (**Table 2**). Using the coalescent simulator, the P-values for Tajima's D could be calculated either with no recombination (very conservative) or intermediate levels of recombination (using estimated recombination from the data sets, R). When the coalescent simulator was used to calculate the P-values for the Tajima's D with no recombination, it became clear that *stwl* was very close to rejecting with a P-value of 0.056, and *ote* did reject the hypothesis of neutrality with a P-value of 0.035. When the coalescent simulator was run with levels of recombination estimated using DnaSP for the two genes (R = 6.8 for *stwl* and R = 36.2 for *ote*), Tajima's D was significant for both *stwl* and *ote* and therefore rejected the hypothesis of neutrality with P-values of 0.035 and 0.005, respectively.

Fay and Wu's H was also calculated (**Table 2**). Fay and Wu's H, like Tajima's

D, also tests how well a data set fits an equilibrium neutral model and thus also assumes a hypothesis of neutrality. Fay and Wu's H was significantly negative for both genes with P-values of 0.007 and 0.013 (obtained from the coalescent simulator run with intermediate levels of recombination with the same estimate of R as noted above) for *stwl* and *ote*, respectively, indicating an excess of high frequency derived mutations.

Table 2 – Population genetics test statistics and P-values for *stwl* and *ote* in *D. pseudoobscura*. All P-values calculated with recombination using the coalescent simulator in DnaSP

	Estimate of R	Tajima's D	P-Value	Fay and Wu's H	P-Value
Stwl no recomb.	6.8	–1.4812	P > 0.10	–7.4689	P = 0.063
Coalescent Simulation – Intermediate Recombination	6.8	–1.4812	P = 0.035	–7.4689	P = 0.007
Ote no recomb.	36.2	–1.6186	0.10 > P > 0.05	–5.4467	P = 0.328
Coalescent Simulation – Intermediate Recombination	36.2	–1.6186	P = 0.005	–5.4467	P = 0.013

However, the McDonald Kreitman Tests, which look for repeated, long-term positive selection, for both *stwl* and *ote* failed to reject the hypothesis of neutrality in *D. pseudoobscura* (**Table 3**). The two-tailed P-value, calculated by Fisher's exact test, was 1.000 for *stwl* and 0.3863 for *ote*.

Table 3 – McDonald Kreitman test for repeated positive selection at *stwl* and *ote* in *D. pseudoobscura*. *P-value calculated by Fisher's exact test (two-tailed) in DnaSP for the McDonald Kreitman Table.

Gene	McDonald Kreitman Table			P – Value*
		Fixed Difference Between Species	Polymorphic Sites	
Stwl	Synonymous Substitutions	26	11	P = 0.61273
	Nonsynonymous Substitutions	33	10	
		Fixed Difference Between Species	Polymorphic Sites	
Ote	Synonymous Substitutions	8	23	P = 0.27489
	Nonsynonymous Substitutions	12	17	

DISCUSSION

For both *stwl* and *ote* the significantly negative Tajima's D and Fay and Wu's H test statistics suggests a significant excess of rare variants, and an excess of high-frequency derived variants, respectively. This excess of rare variants in both genes is consistent with a recent selective sweep at or near these genes has occurred because after a selective sweep the majority of polymorphism comes in the form of singletons. While a major population expansion after a population bottleneck could also produce these same result, Jensen and Bachtrog (2011) found no evidence for such an expansion in their analysis of data from 100 genes across the genome and inferred that *D. pseudoobscura* has maintained a relatively stable size. A selective sweep is, thus, likely to be the cause of the significantly negative Tajima's D and Fay and Wu's H test statistics, than a population expansion. The low synonymous polymorphism, $\pi(s)$, for *stwl* as compared to the average $\pi(s)$ calculated in Jensen and Bachtrog (2011) is also consistent with a recent selective sweep at or very near *stwl*.

In contrast to the results from Tajima's D and Fay and Wu's H for *stwl* and *ote*, which reject the hypothesis of neutrality in a manner suggesting a recent selective sweep,

the McDonald Kreitman tests for both genes fail to reject neutrality, and thus provide no evidence for past recurrent selective fixations favoring new protein variants at either gene. Therefore, because there is no evidence for long term infection of *D. pseudoobscura* with *Wolbachia*, the lack of evidence for long term selection is consistent with the endosymbiont hypothesis. However, there is evidence for a possible recent selective sweep at or near these genes. What drove the putative recent selective sweeps at or near *stwl* and *ote* is currently unclear.

There is evidence that *Wolbachia* both manipulates host reproduction and also can contribute beneficial effects to its host such as resistance to certain types of viruses, which could lend some fitness advantage to flies with the infection (e.g., Hedges et al. 2008; Teixeira, Ferreira, Ashburner 2008). However, *Wolbachia's* manipulation of the host's reproduction causes fertility disadvantages in the uninfected host fly because of cytoplasmic incompatibilities when infected and uninfected flies are crossed (Fry, Palmer, Rand 2004). *Wolbachia* is a candidate for being a driver of positive selection in these GSC genes because of the potential for an "arm's race" between the perceived "advantage" and "disadvantage" of *Wolbachia* in flies (Bauer DuMont et al. 2007). As Bauer DuMont et al. (2007) originally proposed with their endosymbiont conflict hypothesis, there could be a sort of evolutionary balancing act of managing *Wolbachia* infection as to gain its advantages while minimizing its negative effects of reproductive manipulation. Repeated evolution of both the host GSC genes and *Wolbachia* to maximize fitness could lead to the repeated positive selection observed in GSC genes in species with long-term *Wolbachia* infections.

Following the logic of this endosymbiont conflict hypothesis, the results of this

project would suggest several possible scenarios: 1) *D. pseudoobscura* was not infected with *Wolbachia* in the past, but it has become recently infected with *Wolbachia* and evidence of this infection has not been found yet in the relatively limited number of lines (37) of *D. pseudoobscura* that were analyzed by Mateos et al. (2006); 2) *stwl* and *ote* are tightly linked to genes which have undergone recent selective sweeps; or 3) there is another driver of recent positive selection acting on these genes in *D. pseudoobscura*. The results of this project suggest that perhaps more lines of *D. pseudoobscura* should be tested for *Wolbachia* to see if there perhaps has been a recent infection if *D. pseudoobscura* with *Wolbachia*. Alternatively, *D. pseudoobscura* could truly not be infected with *Wolbachia*, and other hypotheses regarding the driver(s) of positive selection of the GSC genes, in at least *D. pseudoobscura,* should be tested. In addition, it would be valuable to assess and analyze sequence variation in the genes that are linked to *stwl* and *ote*, respectively, to attempt to localize the precise target of the recent selective sweeps apparent in those genes.

REFERENCES

Bauer DuMont, V.L., H.A. Flores, M.H. Wright, and C.F. Aquadro, *Recurrent positive selection at bgcn, a key determinant of germ line differentiation, does not appear to be driven by simple coevolution with its partner protein bam.* Mol Biol Evol. **24**(1): p. 182-91, (2007).

Civetta, A., S.A. Rajakumar, B. Brouwers, and J.P. Bacik, *Rapid evolution and gene-specific patterns of selection for three genes of spermatogenesis in Drosophila.* Mol Biol Evol. **23**(3): p. 655-62, (2006).

Fay J.C., Wu C-I. Hitchhiking under positive Darwinian selection. Genetics. 155:1405–1413. (2000).

Flores, H.A., Bauer DuMont, V.L., A. Fatoo, D. Hubbard, M. Hijji, D.A. Barbash, and C.F. Aquadro, *Pervasive adaptive evolution of genes involved in the regulation of Drosophila germline stem cells* (2013).

Fry AJ, Palmer MR, Rand DM. Variable fitness effects of *wolbachia infection in Drosophila melanogaster*. Heredity (Edinb) 93:379-389. (2004).

Gene Codes, Sequencher® version 5.1 sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA http://www.genecodes.com

Hamblin, M.T., C.F. Aquadro, *DNA Sequence Variation and the Recombinational Landscape in Drosophila pseudoobscura: A Study of the Second Chromosome*. The Genetics Society of America. p. 859-869 (1999).

Haselkorn, T.S., T.A. Markow, and N.A. Moran, *Multiple introductions of the Spiroplasma bacterial endosymbiont into Drosophila.* Mol Ecol. **18**(6): p. 1294-305, (2009).

Hedges LM, Brownlie JC, O'Neill SL, Johnson KN. *Wolbachia* and virus protection in insects. Science 322:702. (2008).

Jensen, J.D., D. Bachtrog, *Characterizing the Influence of Effective Population Size on the Rate of Adaptation: Gillespie's Darwin Domain*. Genome Biol. Evol. 3: p. 687–701. doi:10.1093/gbe/evr063. (2011).

Jiang, X., L. Xia, D. Chen, Y. Yang, H. Huang, L. Yang, Q. Zhao, L. Shen, J. Wang, and D. Chen, *Otefin, a nuclear membrane protein, determines the fate of germline stem cells in Drosophila via interaction with Smad complexes.* Dev Cell. 14(4): p. 494-506, (2008).

Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data Bioinformatics 25:1451-1452, (2009).

Marygold, S.J., P.C. Leyland, R.L. Seal, J.L. Goodman, J.R. Thurmond, V.B. Strelets, R.J. Wilson and the FlyBase Consortium, (2013).

Mateos M, Castrezana SJ, Nankivell BJ, Estes AM, Markow TA, Moran NA. Heritable endosymbionts of *Drosophila*. Genetics 174:363-376, (2006).

McDonald JH, Kreitman M. Adaptive protein evolution at the *adh* locus in *Drosophila*. Nature 351:652-654, (1991).

Montenegro, H., V.N. Solferini, L.B. Klaczko, and G.D. Hurst, *Male-killing Spiroplasma natural- ly infecting Drosophila melanogaster.* Insect Mol Biol. **14**(3): p. 281-7, (2005).

Richardson, M.F., L.A. Weinert, J.J. Welch, R.S. Linheiro, M.M. Magwire, F.M. Jiggins, and C.M. Bergman. 2012. Population genomics of the *Wolbachia* endosymbiont in *Drosophila melanogaster*. PLOS Genet 8(12):e1003129.

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism Genetics 123:585-595, (1989).

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0 Mol. Biol. Evol. 24:1596-1599, (2007).

Teixeira L, Ferreira A, Ashburner M. The bacterial symbiont *wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. PLoS Biol. 6:e2. (2008).

Watts, T., T.S. Haselkorn, N.A. Moran, and T.A. Markow, *Variable incidence of Spiroplasma infections in natural populations of Drosophila species.* PLoS One. **4**(5): p. e5703, (2009).

Werren, J.H., L. Baldo, and M.E. Clark, *Wolbachia: Master Manipulators of Invertebrate Biology*. Nature Reviews Microbiology. **6: p.** 741-751, (2008)

Xie, T. *Control of germline stem cell self-renewal and differentiation in the Drosophila ovary: concerted actions of niche signals and intrinsic factors*. *WIREs Dev Biol.* doi: 10.1002/wdev.60. (2012).

APPENDIX A – *STONEWALL* SNP'S

***Stonewall*: Nucleotide Positions of the Polymorphism**
(/ = gap)

	nt 39	nt 128	nt 216	nt 241	nt 271	nt 276	nt 316	nt 353	nt 354
Ancestral	A	C	C	A	A	G	A	T	T
AH130	/	/	.	T	G	A	G	C	G
AH133	/	/	.	T	G	A	G	C	G
AH135	/	/	.	T	G	A	G	C	G
AH144	/	/	.	T	G	A	G	C	G
AH155	/	/	.	T	G	A	G	C	G
AH162	/	/	.	T	G	A	G	C	G
AH172	/	/	.	T	G	A	G	C	G
AH41	/	/	T	T	G	A	G	C	G
BMC10	G	T	.	T	G	A	G	C	G
BMC11	G	T	.	T	G	A	G	C	G
BMC13	G	T	.	T	G	A	G	C	G
BMC3	G	T	.	T	G	A	G	C	G
BMC4	G	T	.	.	G	A	G	C	G
BMC5	G	T	.	T	G	A	G	C	G
BMC7	G	T	.	T	G	A	G	C	G
BMC8	G	T	.	T	G	A	G	C	G
BMC9	G	T	.	T	G	A	G	C	G
KB10	G	T	.	T	G	A	G	C	G
KB12	G	T	.	T	G	A	G	C	G
KB3	G	T	.	.	G	A	G	C	G
KB4	G	T	.	T	G	A	G	C	G
KB5	G	T	.	T	G	A	G	C	G
KB6	G	T	.	.	G	A	G	C	G
KB9	G	T	.	T	G	A	G	C	G
MV1	G	T	.	.	G	A	G	C	G
MV10	G	T	.	T	G	A	G	C	G
MV11	G	T	.	T	G	A	G	C	G
MV18	G	T	.	.	G	A	G	C	G
MV2	G	T	.	T	G	A	G	C	G
MV25	G	T	.	T	G	A	G	C	G
MV26	G	T	.	T	G	A	G	C	G
MV30	G	T	.	.	G	A	G	C	G
MV32	G	T	.	T	G	A	G	C	G
MV6	G	T	.	T	G	A	G	C	G
MV7	G	T	.	T	G	A	G	C	G
'Dpse_stwl	G	T	.	T	G	A	G	C	G
'Mir_tmp'	.	.	C	.	.	.	.	.	.

	nt 465	nt 495	nt 537	nt 566	nt 579	nt 587	nt 590	nt 621	nt 642
Ancestral	G	G	G	G	C	T	A	A	T
AH130	T	A	T	A	A	A	G	G	.
AH133	T	A	T	A	.	A	G	G	.
AH135	T	A	T	A	A	A	G	G	.
AH144	T	A	T	A	A	A	G	G	.
AH155	T	A	T	A	A	A	G	G	.
AH162	T	A	T	A	A	A	G	G	.
AH172	T	A	T	A	A	A	G	G	.
AH41	T	A	T	A	A	A	G	G	.
BMC10	T	A	T	A	A	A	G	G	.
BMC11	T	A	T	A	A	A	G	G	.
BMC13	T	A	T	A	A	A	G	G	.
BMC3	T	A	T	A	.	A	G	G	.
BMC4	T	A	T	A	.	A	G	G	A
BMC5	T	A	T	A	A	A	G	G	.
BMC7	T	A	T	A	A	A	G	G	.
BMC8	T	A	T	A	A	A	G	G	.
BMC9	T	A	T	A	A	A	G	G	.
KB10	T	A	T	A	A	A	G	G	.
KB12	T	A	T	A	A	A	G	G	.
KB3	T	A	T	A	.	A	G	G	.
KB4	T	A	T	A	A	A	G	G	.
KB5	T	A	T	A	A	A	G	G	.
KB6	T	A	T	A	.	A	G	G	.
KB9	T	A	T	A	A	A	G	G	.
MV1	T	A	T	A	.	A	G	G	.
MV10	T	A	T	A	A	A	G	G	.
MV11	T	A	T	A	A	A	G	G	.
MV18	T	A	T	A	.	A	G	G	.
MV2	T	A	T	A	A	A	G	G	.
MV25	T	A	T	A	A	A	G	G	.
MV26	T	A	T	A	A	A	G	G	.
MV30	T	A	T	A	.	A	G	G	.
MV32	T	A	T	A	A	A	G	G	.
MV6	T	A	T	A	A	A	G	G	.
MV7	T	A	T	A	A	A	G	G	.
'Dpse_stwl_	T	A	T	A	A	A	G	G	.
'Mir_tmp'	.	.	.	.	.	.	.	.	.

B

	nt 651	nt 653	nt 663	nt 672	nt 724	nt 784	nt 785	nt 795	nt 834
Ancestral	G	T	G	C	T	T	G	A	C
AH130	A	C	A	.	C	.	.	C	A
AH133	A	C	A	.	C	.	.	C	A
AH135	A	C	A	.	C	.	.	C	A
AH144	A	C	A	.	C	.	.	C	A
AH155	A	C	A	.	C	.	.	C	A
AH162	A	C	A	.	C	.	.	C	A
AH172	A	C	A	.	C	.	.	C	A
AH41	A	C	A	.	C	.	.	C	A
BMC10	A	C	A	.	C	.	.	C	A
BMC11	A	C	A	.	C	.	.	C	A
BMC13	A	C	A	.	C	.	.	C	A
BMC3	A	C	A	.	C	.	.	C	A
BMC4	A	C	A	.	C	G	T	C	A
BMC5	A	C	A	.	C	.	.	C	A
BMC7	A	C	A	.	C	.	.	C	A
BMC8	A	C	A	.	C	.	.	C	A
BMC9	A	C	A	.	C	.	.	C	A
KB10	A	C	A	.	C	.	.	C	A
KB12	A	C	A	.	C	.	.	C	A
KB3	A	C	A	.	C	G	T	C	A
KB4	A	C	A	.	C	.	.	C	A
KB5	A	C	A	.	C	.	.	C	A
KB6	A	C	A	.	C	.	.	C	A
KB9	A	C	A	.	C	.	.	C	A
MV1	A	C	A	T	C	.	.	C	A
MV10	A	C	A	.	C	.	.	C	A
MV11	A	C	A	.	C	.	.	C	A
MV18	A	C	A	.	C	.	.	C	A
MV2	A	C	A	.	C	.	.	C	A
MV25	A	C	A	.	C	.	.	C	A
MV26	A	C	A	.	C	.	.	C	A
MV30	A	C	A	.	C	.	.	C	A
MV32	A	C	A	.	C	.	.	C	A
MV6	A	C	A	.	C	.	.	C	A
MV7	A	C	A	.	C	.	.	C	A
'Dpse_stwl_	A	C	A	.	C	.	.	C	A
'Mir_tmp'	.	.	.	C	.	.	.	.	.

	nt 879	nt 881	nt 911	nt 914	nt 924	nt 927	nt 1005	nt 1020	nt 1068
Ancestral	G	C	G	C	A	C	A	G	G
AH130	.	T	A	G	C	G	C	T	A
AH133	.	T	A	G	C	G	C	T	A
AH135	.	T	A	G	C	G	C	T	A
AH144	.	T	A	G	C	G	C	T	A
AH155	.	T	A	G	C	G	C	T	A
AH162	.	T	A	G	C	G	C	T	A
AH172	.	T	A	G	C	G	C	T	A
AH41	.	T	A	G	C	G	C	T	A
BMC10	.	T	A	G	C	G	C	T	A
BMC11	.	T	A	G	C	G	C	T	A
BMC13	.	T	A	G	C	G	C	T	A
BMC3	.	T	A	G	C	G	C	T	A
BMC4	.	T	A	G	C	G	C	T	A
BMC5	.	T	A	G	C	G	C	T	A
BMC7	.	T	A	G	C	G	C	T	A
BMC8	.	T	A	G	C	G	C	T	A
BMC9	.	T	A	G	C	G	C	T	A
KB10	.	T	A	G	C	G	C	T	A
KB12	.	T	A	G	C	G	C	T	A
KB3	.	T	A	G	C	G	C	T	A
KB4	.	T	A	G	C	G	C	T	A
KB5	.	T	A	G	C	G	C	T	A
KB6	.	T	A	G	C	G	C	T	A
KB9	.	T	A	G	C	G	C	T	A
MV1	.	T	A	G	C	G	C	T	A
MV10	.	T	A	G	C	G	C	T	A
MV11	.	T	A	G	C	G	C	T	A
MV18	.	T	A	G	C	G	C	T	A
MV2	.	T	A	G	C	G	C	T	A
MV25	.	T	A	G	C	G	C	T	A
MV26	A	T	A	G	C	G	C	T	A
MV30	.	T	A	G	C	G	C	T	A
MV32	.	T	A	G	C	G	C	T	A
MV6	.	T	A	G	C	G	C	T	A
MV7	.	T	A	G	C	G	C	T	A
'Dpse_stwl	.	T	A	G	C	G	C	T	A
'Mir_tmp'	.	.	.	.	.	.	.	.	.

	nt 1086	nt 1202	nt 1256	nt 1268	nt 1275	nt 1369	nt 1449	nt 1450	nt 1461
Ancestral	T	C	A	T	A	G	A	G	C
AH130	A	G	G	C	T	.	G	A	T
AH133	A	G	G	C	T	.	G	A	T
AH135	A	G	G	C	T	.	G	A	T
AH144	A	G	G	C	T	A	G	A	T
AH155	A	G	G	C	T	A	G	A	T
AH162	A	G	G	C	T	A	G	A	T
AH172	A	G	G	C	T	A	G	A	T
AH41	A	G	G	C	T	A	G	A	T
BMC10	A	G	G	C	T	A	G	A	T
BMC11	A	G	G	C	T	A	G	A	T
BMC13	A	G	G	C	T	A	G	A	T
BMC3	A	G	G	C	T	.	G	A	T
BMC4	A	G	G	C	T	.	G	A	T
BMC5	A	G	G	C	T	A	G	A	T
BMC7	A	G	G	C	T	.	G	A	T
BMC8	A	G	G	C	T	.	G	A	T
BMC9	A	G	G	C	T	.	G	A	T
KB10	A	G	G	C	T	A	G	A	T
KB12	A	G	G	C	T	.	G	A	T
KB3	A	G	G	C	T	.	G	A	T
KB4	A	G	G	C	T	A	G	A	T
KB5	A	G	G	C	T	.	G	A	T
KB6	A	G	G	C	T	.	G	A	T
KB9	A	G	G	C	T	A	G	A	T
MV1	A	G	G	C	T	A	G	A	T
MV10	A	G	G	C	T	A	G	A	T
MV11	A	G	G	C	T	.	G	A	T
MV18	A	G	G	C	T	.	G	A	T
MV2	A	G	G	C	T	.	G	A	T
MV25	A	G	G	C	T	.	G	A	T
MV26	A	G	G	C	T	A	G	A	T
MV30	A	G	G	C	T	.	G	A	T
MV32	A	G	G	C	T	A	G	A	T
MV6	A	G	G	C	T	.	G	A	T
MV7	A	G	G	C	T	A	G	A	T
'Dpse_stwl_	A	G	G	C	T	.	G	A	T
'Mir_tmp'	.	.	.	.	.	.	.	.	.

	nt 1467	nt 1527	nt 1530	nt 1534	nt 1540	nt 1548	nt 1549	nt 1561	nt 1567
Ancestral	C	A	C	A	C	T	T	C	T
AH130	T	G	T	C	G	C	.	.	A
AH133	T	G	T	C	G	C	.	.	A
AH135	T	G	T	C	G	C	.	.	A
AH144	T	G	T	C	G	C	.	.	A
AH155	T	G	T	C	G	C	.	.	A
AH162	T	G	T	C	G	C	.	.	A
AH172	T	G	T	C	G	C	.	.	A
AH41	.	G	.	C	G	C	.	T	.
BMC10	.	G	.	C	G	C	.	T	.
BMC11	T	G	T	C	G	C	.	.	A
BMC13	T	G	T	C	G	C	.	.	A
BMC3	T	G	T	C	G	C	.	.	A
BMC4	T	G	T	C	G	C	.	.	A
BMC5	T	G	T	C	G	C	.	.	A
BMC7	T	G	T	C	G	C	.	.	A
BMC8	T	G	T	C	G	C	.	.	A
BMC9	T	G	T	C	G	C	.	.	A
KB10	T	G	T	C	G	C	.	.	A
KB12	T	G	T	C	G	C	.	.	A
KB3	T	G	T	C	G	C	.	.	A
KB4	T	G	T	C	G	C	.	.	A
KB5	T	G	T	C	G	C	.	.	A
KB6	T	G	T	C	G	C	.	.	A
KB9	T	G	T	C	G	C	.	.	A
MV1	T	G	T	C	G	C	.	.	A
MV10	T	G	T	C	G	C	.	.	A
MV11	T	G	T	C	G	C	.	.	A
MV18	.	G	T	C	G	C	A	.	A
MV2	T	G	T	C	G	C	.	.	A
MV25	T	G	T	C	G	C	.	.	A
MV26	T	G	T	C	G	C	.	.	A
MV30	T	G	T	C	G	C	.	.	A
MV32	T	G	T	C	G	C	.	.	A
MV6	T	G	T	C	G	C	.	.	A
MV7	T	G	T	C	G	C	.	.	A
'Dpse_stwl	T	G	T	C	G	C	.	.	A
'Mir_tmp'	.	.	.	.	.	.	.	.	.

	nt 1668	nt 1674	nt 1727	nt 1733	nt 1748	nt 1749	nt 1801	nt 1815	nt 1818
Ancestral	C	T	G	A	G	G	A	A	T
AH130	.	G	C	C	A	T	G	T	G
AH133	.	G	C	C	A	T	G	T	G
AH135	.	G	C	C	A	T	G	T	G
AH144	.	G	C	C	A	T	G	T	G
AH155	.	G	C	C	A	T	G	T	G
AH162	.	G	C	C	A	T	G	T	G
AH172	.	G	C	C	A	T	G	T	G
AH41	.	G	C	C	A	T	G	T	G
BMC10	.	G	C	C	A	T	G	T	G
BMC11	.	G	C	C	A	T	G	T	G
BMC13	.	G	C	C	A	T	G	T	G
BMC3	.	G	C	C	A	T	G	T	G
BMC4	G	G	C	C	A	T	G	T	G
BMC5	.	G	C	C	A	T	G	T	G
BMC7	.	G	C	C	A	T	G	T	G
BMC8	.	G	C	C	A	T	G	T	G
BMC9	.	G	C	C	A	T	G	T	G
KB10	.	G	C	C	A	T	G	T	G
KB12	.	G	C	C	A	T	G	T	G
KB3	G	G	C	C	A	T	G	T	G
KB4	.	G	C	C	A	T	G	T	G
KB5	.	G	C	C	A	T	G	T	G
KB6	G	G	C	C	A	T	G	T	G
KB9	.	G	C	C	A	T	G	T	G
MV1	.	G	C	C	A	T	G	T	G
MV10	.	G	C	C	A	T	G	T	G
MV11	.	G	C	C	A	T	G	T	G
MV18	.	G	C	C	A	T	G	T	G
MV2	.	G	C	C	A	T	G	T	G
MV25	.	G	C	C	A	T	G	T	G
MV26	.	G	C	C	A	T	G	T	G
MV30	G	G	C	C	A	T	G	T	G
MV32	.	G	C	C	A	T	G	T	G
MV6	.	G	C	C	A	T	G	T	G
MV7	.	G	C	C	A	T	G	T	G
'Dpse_stwl_	.	G	C	C	A	T	G	T	G
'Mir_tmp'	.	.	.	.	.	.	.	.	.

	nt 1819	nt 1916	nt 1932	nt 2075	nt 2094	nt 2168	nt 2199	nt 2238	nt 2285
Ancestral	T	T	T	C	T	C	A	C	C
AH130	G	C	C	G	G	G	T	.	A
AH133	G	C	C	G	G	G	T	.	A
AH135	G	C	C	G	G	G	T	.	A
AH144	G	C	C	G	G	G	T	T	A
AH155	G	C	C	G	G	G	T	.	A
AH162	G	C	C	/	G	G	T	.	A
AH172	G	C	C	G	G	G	T	.	A
AH41	G	C	C	G	G	G	T	.	A
BMC10	G	C	C	G	G	G	T	.	A
BMC11	G	C	C	G	G	G	T	T	A
BMC13	G	C	C	G	G	G	T	T	A
BMC3	G	C	C	G	G	G	T	.	A
BMC4	G	C	C	G	G	G	T	.	A
BMC5	G	C	C	G	G	G	T	.	A
BMC7	G	C	C	G	G	G	T	.	A
BMC8	G	C	C	G	G	G	T	.	A
BMC9	G	C	C	G	G	G	T	.	A
KB10	G	C	C	G	G	G	T	T	A
KB12	G	C	C	G	G	G	T	.	A
KB3	G	C	C	G	G	G	T	.	A
KB4	G	C	C	G	G	G	T	.	A
KB5	G	C	C	G	G	G	T	.	A
KB6	G	C	C	G	G	G	T	.	A
KB9	G	C	C	G	G	G	T	T	A
MV1	G	C	C	G	G	G	T	.	A
MV10	G	C	C	G	G	G	T	.	A
MV11	G	C	C	G	G	G	T	.	A
MV18	G	C	C	G	G	G	T	.	A
MV2	G	C	C	G	G	G	T	.	A
MV25	G	C	C	G	G	G	T	.	A
MV26	G	C	C	G	G	G	T	.	A
MV30	G	C	C	G	G	G	T	.	A
MV32	G	C	C	G	G	G	T	.	A
MV6	G	C	C	G	G	G	T	.	A
MV7	G	C	C	G	G	G	T	.	A
'Dpse_stwl_	G	C	C	G	G	G	T	.	A
'Mir_tmp'	.	.	.	.	.	.	.	.	.

	nt 2359	nt 2418	nt 2457	nt 2468	nt 2552	nt 2572	nt 2579	nt 2742	nt 2748
Ancestral	A	C	A	C	N	A	A	A	A
AH130	T	T	.	.	C	C	G	G	G
AH133	T	T	.	.	C	C	G	G	G
AH135	T	T	.	.	C	C	G	G	G
AH144	T	T	.	.	C	C	G	G	G
AH155	T	T	.	.	C	C	G	G	G
AH162	T	T	G	.	C	C	G	G	G
AH172	T	T	.	.	C	C	G	G	G
AH41	T	T	.	.	C	C	G	G	G
BMC10	T	T	.	.	C	C	G	G	G
BMC11	T	T	.	.	C	C	G	G	G
BMC13	T	T	.	.	C	C	G	G	G
BMC3	T	T	.	.	C	C	G	G	G
BMC4	T	T	.	.	C	C	G	G	G
BMC5	T	T	.	.	C	C	G	G	G
BMC7	T	T	.	.	C	C	G	G	G
BMC8	T	T	.	.	C	C	G	G	G
BMC9	T	T	.	.	C	C	G	G	G
KB10	T	T	.	.	C	C	G	G	G
KB12	T	T	.	.	C	C	G	G	G
KB3	T	T	.	.	C	C	G	G	G
KB4	T	T	.	.	C	C	G	G	G
KB5	T	T	.	.	C	C	G	G	G
KB6	T	T	.	.	C	C	G	G	G
KB9	T	T	.	.	C	C	G	G	G
MV1	T	T	.	.	T	C	G	G	G
MV10	T	T	.	.	C	C	G	G	G
MV11	T	T	.	.	C	C	G	G	G
MV18	.	T	.	.	C	C	G	G	G
MV2	T	T	.	.	C	C	G	G	G
MV25	T	T	.	.	C	C	G	G	G
MV26	.	T	.	A	C	C	G	G	G
MV30	T	T	.	.	C	C	G	G	G
MV32	T	T	.	.	C	C	G	G	G
MV6	T	T	.	.	C	C	G	G	G
MV7	T	T	.	.	C	C	G	G	G
'Dpse_stwl_	T	T	.	.	C	C	G	G	G
'Mir_tmp'	.	.	.	.	.	.	.	.	.

	nt 2834	nt 2888	nt 2946	nt 2952	nt 2970	nt 2982
Ancestral	A	A	A	T	C	A
AH130	T	G	.	C	.	.
AH133	T	G	G	C	.	.
AH135	T	G	.	C	.	.
AH144	T	G	.	C	.	.
AH155	T	G	.	C	T	T
AH162	T	G	.	C	.	.
AH172	T	G	.	C	.	T
AH41	T	G	.	C	.	.
BMC10	T	G	.	C	.	.
BMC11	T	G	.	C	.	.
BMC13	T	G	.	C	.	.
BMC3	T	G	.	C	.	.
BMC4	T	G	.	C	.	.
BMC5	T	G	.	C	.	/
BMC7	T	G	.	C	.	.
BMC8	T	G	.	C	.	.
BMC9	T	G	.	C	.	.
KB10	T	G	.	C	.	.
KB12	T	G	.	C	.	.
KB3	T	G	.	C	.	.
KB4	T	G	.	C	.	/
KB5	T	G	.	C	.	/
KB6	T	G	.	C	.	.
KB9	T	G	.	C	.	.
MV1	T	G	.	C	.	.
MV10	T	G	.	C	.	/
MV11	T	G	.	C	.	.
MV18	T	G	.	C	.	.
MV2	T	G	.	C	.	.
MV25	T	G	.	C	.	.
MV26	T	G	.	C	.	.
MV30	T	G	.	C	.	.
MV32	T	G	.	C	.	/
MV6	T	G	.	C	.	.
MV7	T	G	.	C	.	/
'Dpse_stwl	T	G	.	C	.	.
'Mir_tmp'	.	.	.	.	.	.

APPENDIX B – *OTEFIN* SNP'S

***Otefin*:** Nucleotide Positions of the Polymorphism
(/ = gap)

	nt 24	nt 161	nt 180	nt 189	nt 196	nt 214	nt 222	nt 223	nt 269
Ancestral	C	C	C	A	C	C	T	G	C
BMC10	/	.	T	.	.	.	C	.	G
BMC11	T	.	T	.	.	.	C	T	G
BMC12	/	.	.	.	.	.	.	.	G
BMC13	/	.	T	.	.	.	C	.	G
BMC3	/	.	.	.	.	.	.	.	G
BMC4	/	.	T	.	.	.	C	.	G
BMC5	/	.	.	.	.	.	.	.	G
BMC7	/	.	.	.	T	.	A	.	G
BMC8	/	.	.	.	.	.	.	.	G
BMC9	T	A	.	G	.	G	.	.	G
KB1	T	.	T	.	.	.	C	.	G
KB3	/	.	.	.	.	.	.	.	G
KB4	T	.	.	.	.	.	.	.	.
KB5	/	.	.	.	.	.	.	.	G
KB6	/	.	T	.	.	.	C	.	G
KB8	/	.	T	.	.	.	C	.	G
MV1	/	.	.	.	.	.	.	.	G
MV10	/	.	.	.	.	.	.	.	G
MV18	/	.	.	.	.	.	.	.	G
MV2	/	.	.	.	.	.	.	.	G
MV24	/	.	.	.	.	.	.	.	G
MV25	/	.	.	.	.	.	.	.	G
MV32	/	.	T	.	.	.	C	.	G
MV6	/	.	T	.	.	.	C	.	G
MV7	/	.	.	.	.	.	.	.	G
'Dpse_ote'	T	.	.	.	.	.	.	.	G
D.miranda|	.	.	.	.	.	.	.	.	.

	nt 271	nt 273	nt 297	nt 327	nt 338	nt 342	nt 385	nt 387	nt 417
Ancestral	G	G	G	C	T	C	A	G	A
BMC10	T	C	.	.	C	.	.	.	G
BMC11	T	C	.	.	C	.	.	.	G
BMC12	T	C	.	.	C	.	C	.	G
BMC13	T	C	.	.	C	.	.	.	G
BMC3	T	C	.	.	C	.	.	.	G
BMC4	T	C	.	.	C	.	.	.	G
BMC5	T	C	.	.	C	.	.	.	G
BMC7	T	C	.	.	C	.	.	.	G
BMC8	T	C	.	.	C	.	.	.	G
BMC9	T	C	.	.	C	.	C	.	G
KB1	T	C	.	.	C	.	.	.	G
KB3	T	C	.	.	C	T	.	.	G
KB4	T	C	.	.	C	.	C	.	G
KB5	T	C	.	.	C	.	.	.	G
KB6	T	C	.	.	C	.	.	.	G
KB8	T	C	.	.	C	.	.	.	G
MV1	T	C	.	.	C	.	.	.	G
MV10	T	C	.	.	C	.	C	.	G
MV18	T	C	.	.	C	T	.	.	G
MV2	T	C	.	.	C	.	.	.	G
MV24	T	C	A	.	C	.	.	.	G
MV25	T	C	.	.	C	.	.	.	G
MV32	T	C	.	A	C	.	.	T	G
MV6	T	C	.	A	C	.	.	T	G
MV7	T	C	A	.	C	.	.	.	G
'Dpse_ote'	T	C	.	.	C	.	.	.	G
D.miranda|	.	.	.	.	.	.	.	.	.

	nt 423	nt 434	nt 438	nt 453	nt 477	nt 489	nt 490	nt 516	nt 546
Ancestral	G	G	A	A	G	A	A	C	A
BMC10	.	.	.	G	.	C	G	.	.
BMC11	.	.	.	G	.	C	G	A	.
BMC12	.	.	C	G	.	C	G	.	.
BMC13	.	.	.	G	.	C	G	.	.
BMC3	.	.	.	G	.	C	G	.	.
BMC4	.	.	.	G	.	C	G	.	.
BMC5	.	.	.	G	T	C	G	.	T
BMC7	.	.	.	G	.	C	G	.	.
BMC8	.	.	.	G	.	C	G	.	.
BMC9	.	.	.	G	.	C	G	.	.
KB1	G	.	.	G	.	C	G	A	.
KB3	.	C	.	G	.	C	G	.	.
KB4	.	.	.	G	.	C	G	.	.
KB5	.	.	.	G	.	C	G	.	.
KB6	.	.	.	G	.	C	G	.	.
KB8	.	.	.	G	.	C	G	A	.
MV1	.	.	.	G	.	C	G	.	.
MV10	.	.	.	G	.	C	G	.	.
MV18	.	.	.	G	.	C	G	.	.
MV2	.	.	.	G	.	C	G	.	.
MV24	.	.	.	G	.	C	G	.	.
MV25	A	.	.	G	.	C	G	.	.
MV32	.	.	.	G	.	C	G	.	.
MV6	.	.	.	G	.	C	G	.	.
MV7	.	.	.	G	.	C	G	.	.
'Dpse_ote'	.	.	.	G	.	C	G	.	.
D.miranda|	.	.	.	A	.	.	.	.	.

	nt 570	nt 595	nt 609	nt 645	nt 656	nt 679	nt 698	nt 704	nt 712
Ancestral	C	A	A	A	A	G	T	G	G
BMC10	T	T	.	G	T	T	.	C	.
BMC11	T	T	.	G	T	T	.	C	.
BMC12	T	T	.	G	T	T	.	C	.
BMC13	T	T	T	G	T	T	.	C	.
BMC3	T	T	.	G	T	T	.	C	.
BMC4	T	T	.	G	T	T	.	C	.
BMC5	T	.	.	G	T	T	C	C	.
BMC7	T	T	.	G	T	T	.	C	.
BMC8	T	T	.	G	T	T	.	C	.
BMC9	T	T	.	G	T	T	.	C	.
KB1	T	T	.	G	T	T	.	C	.
KB3	T	T	.	G	T	T	.	C	.
KB4	T	T	.	G	T	T	.	C	.
KB5	T	T	.	G	T	T	.	C	.
KB6	T	T	.	G	T	T	.	C	.
KB8	T	T	.	G	T	T	.	C	.
MV1	T	T	.	G	T	T	.	C	.
MV10	T	T	.	G	T	T	.	C	.
MV18	T	T	.	G	T	T	.	C	.
MV2	T	T	.	G	T	T	.	C	.
MV24	T	T	.	G	T	T	.	C	.
MV25	T	T	.	G	T	T	.	C	.
MV32	T	T	.	G	T	T	.	C	.
MV6	T	T	.	G	T	T	.	C	.
MV7	T	T	.	G	T	T	.	C	A
'Dpse_ote'	T	T	.	G	T	T	.	C	.
D.miranda|	.	.	.	.	.	.	.	.	.

	nt 716	nt 717	nt 738	nt 754	nt 780	nt 795	nt 918	nt 945	nt 946
Ancestral	T	G	A	C	T	A	C	G	C
BMC10	C	.	.	.	G	G	.	.	.
BMC11	C	.	.	.	G	G	.	.	.
BMC12	C	.	G	.	G	.	.	.	.
BMC13	C	.	.	.	G	G	.	.	.
BMC3	C	.	.	.	G	G	.	A	.
BMC4	C	.	.	.	G	G	.	A	.
BMC5	C	.	G	.	G	.	G	.	.
BMC7	C	.	.	.	G	G	.	A	.
BMC8	C	.	.	.	G	T	.	.	.
BMC9	C	.	G	.	G	.	.	.	A
KB1	C	.	.	.	G	G	.	.	.
KB3	C	.	.	.	G	G	.	.	.
KB4	C	.	G	.	G	.	.	.	.
KB5	C	.	.	.	G	G	.	A	.
KB6	C	.	.	.	G	G	.	.	.
KB8	C	.	.	.	G	G	.	.	.
MV1	C	.	.	.	G	G	.	A	.
MV10	C	.	G	.	G	.	.	.	.
MV18	C	.	.	.	G	G	.	.	.
MV2	C	.	.	.	G	T	.	.	.
MV24	C	.	.	.	G	G	.	.	.
MV25	C	.	.	.	G	T	.	.	.
MV32	C	A	.	.	G	.	.	A	.
MV6	C	.	.	.	G	G	.	.	.
MV7	C	.	.	.	G	G	.	.	.
'Dpse_ote'	C	.	.	A	G	G	.	A	.
D.miranda|	.	.	.	.	.	.	.	.	.

	nt 963	nt 965	nt 1026	nt 1029	nt 1090	nt 1098	nt 1104	nt 1113	nt 1120
Ancestral	G	C	G	C	T	A	C	C	A
BMC10	C	.	A	.	C	.	T	A	G
BMC11	C	.	A	G	C	.	T	A	G
BMC12	C	.	A	.	C	C	T	A	G
BMC13	C	.	A	.	C	.	T	A	G
BMC3	T	.	A	.	C	.	T	A	G
BMC4	T	.	A	.	C	.	T	A	G
BMC5	C	.	.	.	C	.	T	A	G
BMC7	T	.	A	.	C	.	T	A	G
BMC8	C	.	.	.	C	.	T	A	G
BMC9	C	.	A	.	C	C	T	A	G
KB1	C	.	A	.	C	.	T	A	G
KB3	C	.	A	.	C	.	.	A	G
KB4	T	.	A	.	C	.	T	A	G
KB5	T	.	A	.	C	.	T	A	G
KB6	C	.	A	.	C	.	T	A	G
KB8	C	.	A	.	C	.	T	A	G
MV1	T	.	A	.	C	.	T	A	G
MV10	C	.	A	.	C	C	T	A	G
MV18	C	.	A	.	C	.	T	A	G
MV2	C	T	A	.	C	.	T	A	G
MV24	C	.	A	.	C	.	T	A	G
MV25	C	.	A	.	C	.	T	A	G
MV32	T	.	A	.	C	.	T	A	G
MV6	C	.	A	.	C	.	T	A	G
MV7	C	.	A	.	C	.	T	A	G
'Dpse_ote'	T	.	A	.	C	.	T	A	G
D.miranda|	.	.	.	.	.	.	.	.	.

	nt 1123	nt 1130	nt 1164	nt 1182	nt 1191	nt 1212
Ancestral	G	T	G	A	A	C
BMC10	A	.	.	T	T	T
BMC11	A	.	.	T	T	T
BMC12	A	.	.	T	T	/
BMC13	A	.	.	T	T	/
BMC3	A	.	.	T	T	T
BMC4	A	.	.	T	T	T
BMC5	A	C	A	T	T	T
BMC7	A	.	.	T	T	T
BMC8	A	.	.	T	T	T
BMC9	A	.	.	T	T	T
KB1	A	.	.	T	T	T
KB3	A	.	.	T	.	T
KB4	A	.	.	T	T	T
KB5	A	.	.	T	T	T
KB6	A	.	.	T	T	T
KB8	A	.	.	T	T	T
MV1	A	.	.	T	T	/
MV10	A	.	.	T	T	T
MV18	A	.	.	T	T	T
MV2	A	.	.	T	T	T
MV24	A	.	.	T	T	T
MV25	A	.	.	T	T	T
MV32	A	.	.	T	T	T
MV6	A	.	.	T	T	T
MV7	A	.	.	T	T	T
'Dpse_ote'	A	.	.	T	T	T
D.miranda|	.	.	.	.	.	.

APPENDIX C – FLY LINES

Duplicate fly lines used to check integrity of the sequence and quality of the sequence editing

Fly lines successfully sequenced:

Stwl: (The larger second exon of *stwl* was sequenced in two segments because of its size)

Segment 2:	Segment 3:
BMC: 7, 8, 9, 11	BMC: 3, 7, 8, 9, 11 10, 12, 13
KB: 3, 5, 8, 9, 12	KB: 1, 3, 4, 5, 6, 9, 10
MV: 1, 7, 10, 11, 18, 26, 28	MV: 7, 11, 25, 26, 32

Ote:

BMC: 3, 4, 5, 7, 8, 9, 10, 12, 13

KB: 1, 3, 4, 3, 6, 8, 9, 10, 12,

MV: 1, 2, 6, 7, 10, 11, 18, 30, 32

Supplemental fly lines sequenced by Jae Choi (Graduate Student):

Stwl:

Segment 1 (~100 base pair first exon), Segment 2, and Segment 3:

AH: 41, 130, 133, 135, 144, 155, 162, 172

BMC: 3, 4, 5, 7, 8, 9, 10, 11, 13

KB: 3, 4, 5, 6, 9, 10, 12

MV: 1, 2, 6, 7, 10, 11, 18, 25, 26, 30, 32

Ote:

BMC: 3, 4, 5, 7, 8, 9, 10, 11, 12, 13

KB: 1, 3, 4, 5, 6, 8

MV: 1, 2, 6, 7, 10, 18, 24, 25, 32